

Polarization as Texture: Microscale 3D Shape from Polarized Light Focus

Ren Matsumoto
Kyushu Institute of Technology*
matsumoto.ren209@mail.kyutech.jp

Takahiro Okabe
Okayama University†
okabe@okayama-u.ac.jp

Ryo Kawahara
Kyoto University‡
ryo@vision.ist.i.kyoto-u.ac.jp

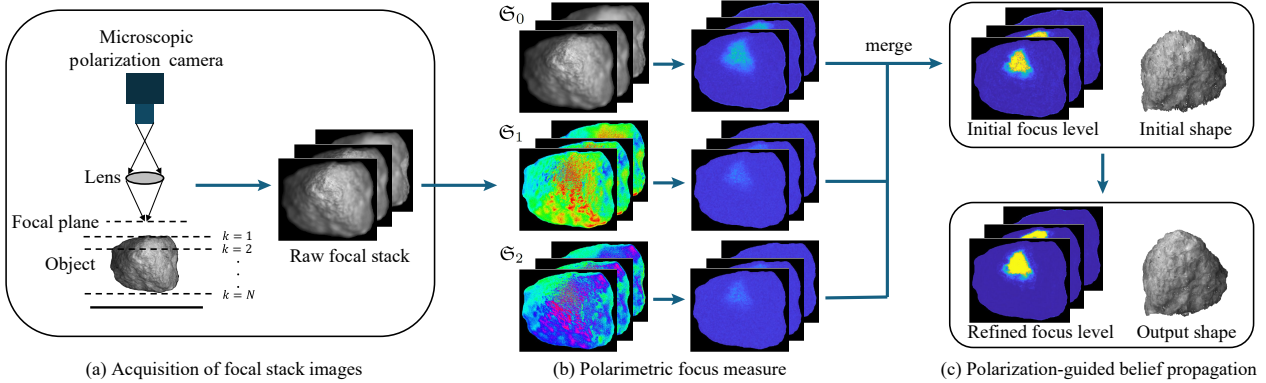


Figure 1. Shape from Polarized Light Focus. We show that (a) capturing different depth-focused stack images with a microscope polarization camera provides (b) effective textures for focus level estimation. (c) Our method further utilizes normal cues from polarization and recovers per-pixel accurate shapes of tiny objects.

Abstract

Defocus is a crucial cue for image-based microscale depth estimation, yet its measurement depends on spatial appearance changes, such as texture. We show that passively observed polarization is responsive to small irregularities of the surface visible in the microscopic world and can be leveraged for focus measure as a strong texture. Our key idea is to leverage texture from polarization for blur analysis and accurately estimate the focus level of microscopic polarization images. We further utilize normal cues from polarization to create a prior distribution of the focus level between neighboring pixels. We then interpolatively propagate the focus level of discrete image slices at different focus depths while denoising. We implement our method with a single polarization camera with a microscope and recover the per-pixel depth from the multi-focus images. The reconstructed results demonstrate the effectiveness of our method for various microscale objects regardless of the surface texture.

1. Introduction

Microscale shape recovery has profound applications in medical and bioinformatics, such as the non-contact and non-invasive analysis of insects and fertilized eggs, the inspection of precision instruments, and material science applications. A common problem in microscopic capturing is blur caused by shallow depth of field, making applying many successful methods and camera setups impractical.

Shape from Focus (SfF) [15, 32] is an innovative passive shape recovery method that leverages this blur’s depth dependence. Suppose we capture multiple images while moving the focus depth. The depth is then estimated using the fact that the most in-focus position corresponds to the depth for each pixel. However, SfF is limited by its reliance on spatial variations in surface appearance, such as texture, to identify blur. Therefore, applying SfF to textureless objects is still an open problem.

In a visual world that differs from the scale we humans see in our daily lives, are textures the only things that cause spatial changes in appearance? In the close-up world, there is often irregularities on surfaces we do not notice, and these shape changes can show new patterns in the reflection of light. If there is an efficient way to capture such small changes, it can improve blur analysis accuracy, and also provide clues to the fine surface geometry.

*Department of Artificial Intelligence, Kyushu Institute of Technology, 680-4 Kawazu, Iizuka, Fukuoka 820-8502, Japan

†Information Technology Track, Faculty of Engineering, Okayama University, 3-1-1 Tsushima-naka, Kita-ku, Okayama 700-8530, Japan

‡Graduate School of Informatics, Kyoto University, Yoshida-honmachi, Sakyo-ku, Kyoto, 606-8501, Japan

In this paper, we show that polarization behaves like texture that depends on the surface orientation and recovers accurate per-pixel depth from multi-focus images. Our key idea is to leverage texture from polarization for analyzing the image’s focus blur and accurately estimate the focus level of microscopic polarization images. Additionally, to effectively exploit the normal cues from polarization, we determine the global shape by designing a prior distribution of focus levels between adjacent pixels based on observed polarization. For this, we filter the focus-level volume by belief propagation with our joint probability designed from the normal cue.

We implement our method with a passive imaging system consisting of a polarization camera, a microscope, and a coaxial-falloff illumination for diffuse polarization analysis. We quantitatively validate the effectiveness of our polarimetric focus measure and polarization-guided belief propagation through evaluation and ablation experiments using synthetic images. We also evaluate the effectiveness of our method qualitatively on various real-world objects. The results clearly demonstrate that our method can recover the accurate 3D shape of microscale objects regardless of surface texture.

Our major contributions are as follows:

- A novel SfF method utilizing polarization for per-pixel depth recovery of tiny textured and textureless objects.
- Derivation of a blurring model in polarization images and a novel focus measure method using Stokes vectors based on this model.
- A novel method of updating global focus level based on belief propagation leveraging normal cues from polarization.

2. Related Work

Most image-based 3D shape reconstruction methods face challenges when applied to microscale objects. For instance, multiview stereo [16] requires precise physical alignment and handling of the blur, and in photometric stereo [26, 46], the normals are affected by blur and smoothed out. As another approach, Interferometric techniques [23, 25] or differential phase contrast techniques [9, 42] are effective in achieving high-depth resolution. Unfortunately, many of these methods require active light sources, or even when natural light is utilized, the complex optical system increases construction costs and form factors.

Shape from Focus and Defocus Shape from focus and defocus has been proposed as a method for 3D shape recovery using focal blur as a depth cue rather than a disturbance [3, 15, 32, 34]. In SfF, the accuracy of measuring focus

level is a critical factor in determining focus depth. Focus level is closely related to local sharpness and is also used for autofocus [14, 48] and acquiring all-in-focus images [1, 27].

Numerous approaches [35] have been proposed to extract the focus level from the image, including gradient-based [2, 29], Laplacian-based [4, 32, 40], and wavelet-based [17, 47, 48] operators. However, they are theoretically based on changes in spatial appearance or features in a window and thus become unreliable for textureless surfaces. Surh *et al.* [38] proposed a ring difference filter with gaps to remove neighboring pixels and calculated semi-local focus levels. Nevertheless, determining the appropriate gap size depends on the object, and accuracy is limited for large textureless areas.

Shape from Polarization In the context of shape cues in passive sensing, polarization can provide normal cues. When light is reflected by an object’s surface, the reflected light is polarized depending on the surface normal. Shape from Polarization (SfP) method utilizes this property to recover surface normals, but as an inverse problem, it is known that ambiguities exist in normal estimation [5, 31].

Therefore, resolving this ambiguity is one of the core issues of SfP. Numerous SfP methods have been proposed, combined with multiview [11, 49–51], shading [19, 33, 36, 39, 43, 44], active lighting [21, 28], shape assumptions such as surface integrability, and deep prior [6, 12, 20, 24]. However, most of these techniques are not directly applicable to microscale objects and are influenced by blurring.

The concept of regarding polarization as texture has been proposed by a few of the binocular stereo [18, 41]. However, as the observed polarization is view-dependent [7, 13], encoders that extract view-independent features must be trained to utilize these characteristics effectively. Additionally, as a stereo method, accuracy is also influenced by the texture of the object’s surface.

In contrast to the above, we utilize polarization as a texture to SfF without changing the viewing direction. We derive the blurring model of polarization and analyze it for shape reconstruction. In addition, we propose a method to design a global shape prior to distribution free from the ambiguity of surface normal in the observed polarization, inspired by the concept of iso-depth contour [11].

3. Background: Polarimetric Defocus Model

We first briefly introduce polarization and describe how polarized light is blurred for the focus analysis.

3.1. Polarization

When light travels as a wave, it oscillates perpendicular to the direction of travel. Unpolarized light, such as sunlight or incandescent light, consists of oscillations in random directions perpendicular to the direction of travel. In contrast,

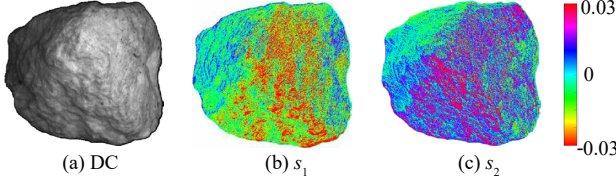


Figure 2. Stokes image reveals spatial changes in surface orientation. (b,c) The s_1 and s_2 components efficiently respond to the normal azimuth's orientations and can extract features different from (a) the DC component's appearance change.

polarized light has a biased direction of oscillation. To examine the degree and orientation of such bias, we can utilize a linear polarizer and extract a specific angular component of polarized light, as employed in polarization cameras.

Concerning the light observed by a polarization camera, observed intensity $I(\alpha)$ for the filter angle α of the linear polarizer is described as

$$I(\alpha) = \bar{I} + \rho \bar{I} \cos(2\alpha - 2\phi), \quad (1)$$

where ρ represents the degree of polarization (DoP), and ϕ represents the angle of polarization (AoP). DoP is the amplitude ratio to the average component $\bar{I}$, and AoP is the filter angle such that $I(\alpha = \phi, \phi + \pi)$ reaches its maximum value. Note that in the case of circularly polarized light, the plane of oscillation also rotates as the light propagates. In this paper, we only consider the change in the dominant linear polarization since we observe the reflections of unpolarized light sources.

Recent division-of-focal-plane (DoFP) polarization cameras, which utilize a quad-Bayer for linear polarizers, can capture polarization images with four filter angles $\alpha = (0, \pi/4, \pi/2, 3\pi/4)$ from a single image. It allows for the linear estimation of the parameters in Eq. 1 for each pixel [19]. For blur analysis, our method especially focuses on the Stokes vector $S = (s_0, s_1, s_2, s_3)^\top$, computed only by adding and subtracting each filter angle intensity;

$$S = \begin{pmatrix} I(0) + I(\frac{\pi}{2}) \\ I(0) - I(\frac{\pi}{2}) \\ I(\frac{\pi}{4}) - I(\frac{3\pi}{4}) \\ 0 \end{pmatrix} = \begin{pmatrix} 2\bar{I} \\ 2\rho\bar{I}\cos 2\phi \\ 2\rho\bar{I}\sin 2\phi \\ 0 \end{pmatrix}. \quad (2)$$

3.2. Observation of Object Surface

Dichromatic reflection model When unpolarized light is reflected on a dielectric surface, the polarization state changes depending on the Fresnel effect, resulting in partially polarized light. The DoP and AoP of reflected light depend on the normal of the object's surface, and their characteristics differ between specular and diffuse reflections. In this study, following the pBRDF [7], we rewrite Eq. 1 at

a certain wavelength by the dichromatic reflection model to exploit the normal cues;

$$I(\alpha) = \bar{I}_s + \rho_s \bar{I}_s \cos(2\alpha - 2\phi_s) + \bar{I}_d + \rho_d \bar{I}_d \cos(2\alpha - 2\phi_d), \quad (3)$$

SfP methods typically reconstruct the shape by utilizing the fact that $\rho_s, \rho_d, \phi_s, \phi_d$ are a function of the surface normal. However, the complexity arises from separating specular and diffuse components and the microfacet normal distribution for specular reflection, which hinders accurate recovery. Instead of analyzing Eq. 3 for arbitrary lighting environments, we employ passive measurements with coaxial falloff lighting to simplify the model.

Polarization Measurement Under Coaxial Falloff Illumination

We employ lighting such that only the diffuse reflection is partially polarized. When using coaxial falloff illumination, the DoP of specular reflection ρ_s becomes 0 because the angle of incidence is 0° . Note that the DoP of the specular lobe, which is caused by microfacets, also becomes 0. Hence, observed intensity of Eq. 3 becomes

$$I(\alpha) = (\bar{I}_s + \bar{I}_d) + \rho_d \bar{I}_d \cos(2\alpha - 2\phi_d). \quad (4)$$

The representation of diffuse polarization is effectively modeled as the transmission of unpolarized light from the object's surface layer, which depends on the principal normal of the point of interest [7]. That is, for the surface normal denoted by $\mathbf{n} = (\cos \psi \sin \theta, \sin \psi \sin \theta, \cos \theta)^\top$, the following hold;

$$\rho_d = \frac{T_{\parallel}(\theta) - T_{\perp}(\theta)}{T_{\parallel}(\theta) + T_{\perp}(\theta)}, \quad \psi = \phi_d \text{ or } \phi_d + \pi, \quad (5)$$

where $T_{\parallel}(\theta)$, $T_{\perp}(\theta)$ denote Fresnel transmittance of p-polarized and s-polarized components at the observing wavelength.

Fig. 2 shows s_1 and s_2 components calculated from polarization images of a *Zeolite* as an example. It clearly shows that the polarization behaves like a texture due to the normal changes on the surface.

3.3. Blurring Model

Since DoFP polarization camera has linear polarizers on the sensor side, the effect of focus blur is modeled as the spread of the light just before it passes through the linear polarizers. Let us denote the depth-dependent blur kernel as $h(z_p)$ and blurred Stokes vector as S' . We can describe the whole process by Muller calculus as follows,

$$S'_\alpha = L(\alpha)S' = L(\alpha) \left\{ \begin{pmatrix} 2(\bar{I}_s + \bar{I}_d) \\ 2\rho_d \bar{I}_d \cos 2\phi_d \\ 2\rho_d \bar{I}_d \sin 2\phi_d \\ 0 \end{pmatrix} * h(z_p) \right\}, \quad (6)$$

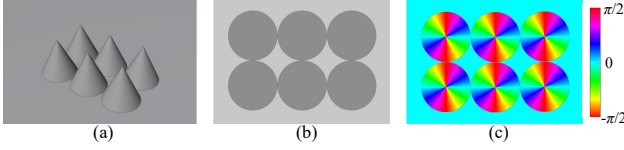


Figure 3. Shape-dependent spatially varying appearance. When we choose a viewpoint under coaxial falloff illumination, a surface of the shape that has an appearance in (a) provides shading as in (b). On the other hand, polarization can reveal different shape features, with the angle of polarization appearing as in (c).

where z_p denotes the depth at pixel p , and $*$ denotes the operator of the convolution for the image domain. $L(\alpha)$ is the following Muller matrix representing the polarizer [10],

$$L(\alpha) = \frac{1}{2} \begin{pmatrix} 1 & \cos 2\alpha & \sin 2\alpha & 0 \\ \cos 2\alpha & \cos^2 2\alpha & \cos 2\alpha \sin 2\alpha & 0 \\ \sin 2\alpha & \cos 2\alpha \sin 2\alpha & \sin^2 2\alpha & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \quad (7)$$

Up to this point, the observed polarization image with focal blur can be described as

$$\begin{aligned} I'(\alpha) &= s'_{\alpha 0} \\ &= (I_s + I_d) * h(z_p) + \rho_d \bar{I}_d \cos(2\alpha - 2\phi_d) * h(z_p) \\ &= I(\alpha) * h(z_p). \end{aligned} \quad (8)$$

Eq. 8 indicates that we can obtain a blurred Stokes vector S' in the same manner as Eq. 2. Also, we can extend this blurring model to the division-of-time polarization cameras, which typically have linear polarizers in front of the camera lens.

Shading feature v.s. AoP feature Even when polarization cannot be observed, shading also works as a spatially varying pattern that depends on shape. To intuitively understand the difference between the two and the usefulness of polarization, let us consider the case when an object with the shape shown in Fig. 3(a) is observed under coaxial falloff illumination. The spatial variation of radiance due to shading depends on the angle between the surface normal and the light source direction. This means the radiance is not affected by normal changes around the light source direction vector, and this insensitivity is visible in some instances, as shown in Fig. 3(b). In contrast, diffuse polarization depends not only on the angle between the surface normal and the viewing direction, but also on the change around the viewing direction vector. For instance, AoP of diffuse polarization appears as shown in Fig. 3(c).

4. Proposed Method

We leverage polarization with the polarimetric focus measure and polarization-guided belief propagation for the focus level estimation and reconstruct an accurate shape per pixel. Our method uses a few realistic assumptions, as follows;

- Orthographic projection for the camera
- Coaxial falloff illumination (a directional light source that has the same direction as the viewing direction)
- Dielectric surface for the objects

4.1. Polarimetric Focus Measure

In this section, we describe how the proposed method measures the focus level of each pixel in polarization images and identifies the focus position in 3D space.

Acquisition of polarization focal stack We utilize a monochrome DoFP polarization camera capable of simultaneously acquiring polarization information in four directions $(0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4})$ to capture the polarization images. As illustrated in Fig. 1(a), the focal stack of raw images is obtained by capturing N steps while sweeping the focal point at known constant intervals along the depth direction. The captured stack is then converted to a polarization focal stack $\mathcal{J}_\alpha \in \mathbb{R}^{H \times W \times N}$ ($\alpha = 0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}$) by a demosaicing process [30].

Focus measure from Stokes vector Intuitively, depending on the object's surface depth, each pixel p will be best focused at a sub-frame of the N image stacks. To detect this, we need a measure of how much is in focus, and measuring the focus level of each image pixel is called focus measure (FM). We follow the survey by Pertuz *et al.* [35] and employ the wavelet-based method [48] for the FM. That is, for each single image, FM evaluates the focus level for each pixel p as

$$f_p = \sum_{p \in \Omega_D} (|W_{LH}(p)| + |W_{HL}(p)| + |W_{HH}(p)|), \quad (9)$$

where Ω_D denotes the local window. W_{LH} , W_{HL} , and W_{HH} denote the vertical, horizontal, and diagonal high-frequency components from the discrete wavelet transform, respectively.

We apply the FM to the Stokes vector representation of the images described in Sec. 3.3 to efficiently evaluate the focus levels of the unpolarized and polarized components. To this end, we convert the polarization focal stack $\mathcal{J}_\alpha$ using Eq. 2 to obtain a focal stack of Stokes images. Let us denote this stack of images as element-wise stacks $\mathcal{S}_0, \mathcal{S}_1, \mathcal{S}_2 \in \mathbb{R}^{H \times W \times N}$, ignoring the s_3 component (Fig. 1(b)). By utilizing the FM in Eq. 9, we can then extract the focus levels

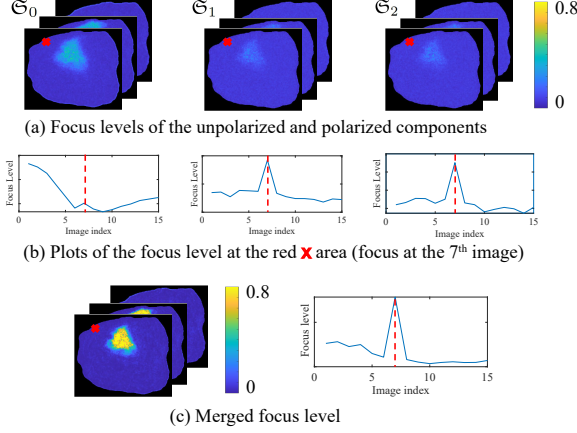


Figure 4. The strategy of polarimetric focus measure based on the Stokes vector aims to capture (a, b) the focus level of radiance as well as the focus level of polarization and (c) merging these to suppress false peaks.

from $\mathfrak{S}_0$, $\mathfrak{S}_1$, and $\mathfrak{S}_2$ for the DC, horizontally polarized, and 45° polarized components.

Ideally, the focus level from FM should have a single peak for the sweep direction for each pixel. However, false peaks can be measured due to depth distribution around the pixel of interest and noise effects in general. Fig. 4 shows an overview of the polarization FM calculation to suppress false peaks. Fig. 4(a) shows an example of the focus level calculation for stacks of $N = 15$ Stokes element images. Picking up the pixel at the red $\times$ position, the distribution of normalized focus level in the depth direction becomes each of Fig. 4(b). For $\mathfrak{S}_0$, $\mathfrak{S}_1$, and $\mathfrak{S}_2$ alone, the focus level is not clear near the peak position, and the focus level has a certain magnitude even far from in-focus depth.

Therefore, as shown in Fig. 4(c), we merge each element to suppress false peaks while enhancing common regions of high focus levels, which contributes to robust peak detection. As a simple approach as a joint probability, we multiply each focus level and normalize it to obtain the distribution $F \in \mathbb{R}^{H \times W \times N}$ of focus levels.

4.2. Polarization-guided Belief Propagation

The FM above measures focus levels locally. To improve the accuracy of the focus level volume, we utilize shape cues from polarization to guide the global distribution of focus locations in 3D space. Inspired by the concept of iso-depth contour tracing with polarization [11], we filter the focus level volume by graph optimization.

Consider a Markov random field (MRF) with each pixel as a node. We assign a random variable f_p to represent the focus level for each node p , and let $\mathbf{f}$ be the set of all f_p for all nodes. When the true focus level $\hat{\mathbf{f}}$ is observed as $\mathbf{f}$ by the FM of the captured stack, our goal is to maximize the

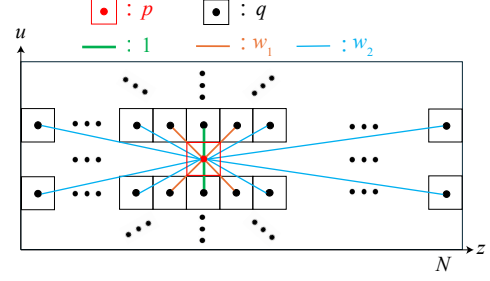


Figure 5. Overview sketch of joint weights for our belief propagation. The pixel p propagates the belief to the adjacent pixel q based on a joint weight designed from the normal cue of polarization.

following posterior probability.

$$P(\hat{\mathbf{f}}|\mathbf{f}) \propto P(\mathbf{f}|\hat{\mathbf{f}})P(\mathbf{f}). \quad (10)$$

Note that $P(\mathbf{f}|\hat{\mathbf{f}})$ is obtained from the measured focus level of observation. Let us consider modeling prior $P(\mathbf{f})$ by an MRF. Since the focus position is shape-dependent, a joint probability $\varphi_{pq}(f_p, f_q)$ can be defined for the focus level for neighboring pixels. The MRF is then described as the joint probability for all links of nodes as follows,

$$P(\mathbf{f}) = \frac{1}{Z} \prod_p \prod_{q \in \mathcal{N}(p)} \varphi_{pq}(f_p, f_q), \quad (11)$$

where Z is a normalization constant and $\mathcal{N}(p)$ denotes the set of neighboring nodes. Specifically, when considering two neighboring pixels p, q as shown in Fig. 5, the random variable f_p has links with f_q for N depths. Therefore, $\varphi_{pq}(f_p, f_q)$ can be determined by the following function $\Phi(z_p, p, q)$;

$$\Phi(z_p, p, q) = \begin{cases} 1/Z_\Phi & (z_p = z_q), \\ w_1/Z_\Phi & (|z_p - z_q| = \Delta d), \\ w_2/Z_\Phi & \text{otherwise,} \end{cases} \quad (12)$$

where z_p and z_q represent the depths of pixels p and q , respectively. Δd is the constant distance between the focal planes of two consecutive stacks, and Z_Φ is the normalization constant for N depths. The weights w_1 and w_2 satisfy $1 \geq w_1 \geq w_2 \geq 0$. Note that the weights are uniformly set to w_2 when the $|z_p - z_q|$ value is larger than Δd to address discontinuous edges.

Our method calculates the values of w_1 and w_2 utilizing the normal azimuth angle obtained from polarization. As described in Eq. 5, the normal azimuth has two potential candidates $(\phi_d, \phi_d + \pi)$ for the observation. Since these two candidates limit the potential gradient direction, the contour is expected to align along the azimuth angle $\phi_d + \frac{\pi}{2}$ or $\phi_d + \frac{3\pi}{2}$ [11]. Therefore, in order to express the joint probability with 8 neighbors for each pixel, our proposed method sets

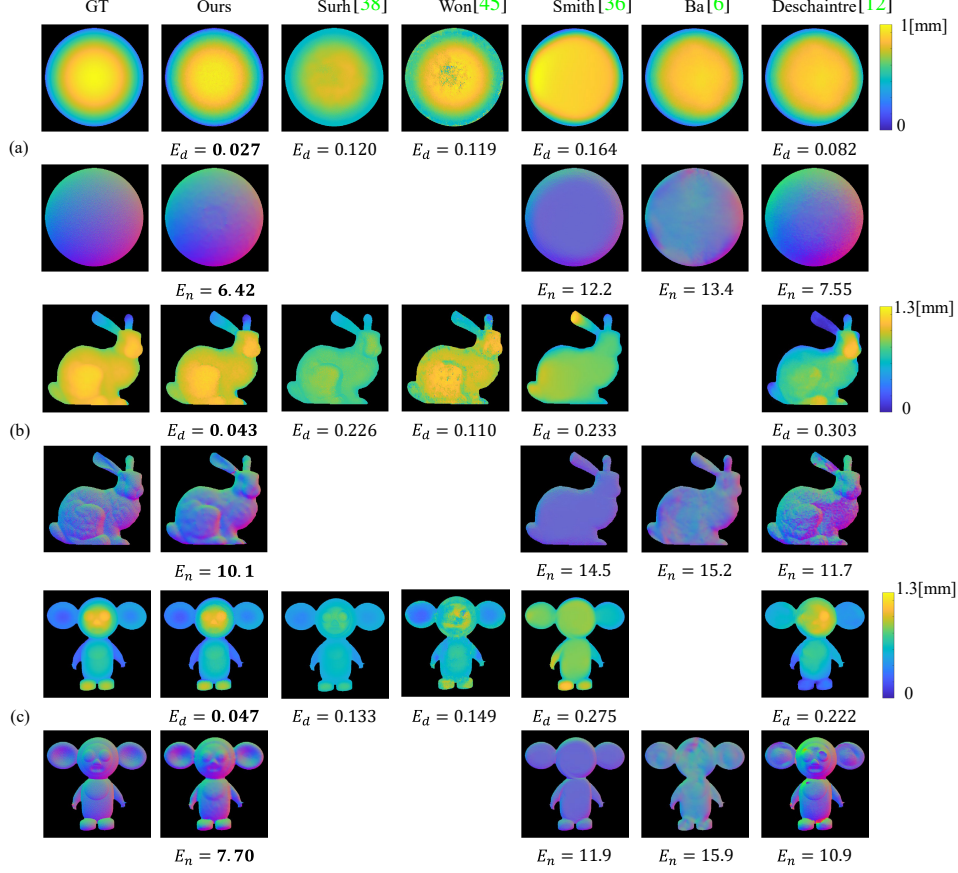


Figure 6. Depth evaluation with synthetic data. The size of (a,b,c) *Sphere* (pBRDF:White biliard), *Bunny* (pBRDF:PEEK) and *Cheburashka* (pBRDF:ZrO2) are approximately 1mm. The blank area is because the corresponding method does not output the depths or the normals. E_d , E_n below the images denote the RMSE in mm and the MAE in degree, respectively.

small w_1 and w_2 to emphasize the weight at $z_p = z_q$ when p and q are aligned in the contour direction (pixels with an azimuth of $0^\circ \pm 30^\circ$ as horizontal, those with an azimuth of $45^\circ \pm 15^\circ$ or $-45^\circ \pm 15^\circ$ as diagonal, and those with an azimuth of $90^\circ \pm 30^\circ$ as vertical.) and sets more uniform weights in other cases. Note that the DoP in Eq. 5 is not used here. This is because the DoP for diffuse polarization is sensitive to noise, and by utilizing only the AoP as a normal cue, we avoid refractive index estimation and dealing with the specular component in Eq. 4.

We apply belief propagation (BP) to the maximization problem above and propagate beliefs until convergence. Let $\hat{F} \in \mathbb{R}^{H \times W \times N}$ represent the focus level obtained by the BP. The final depth is determined by estimating the peak based on discrete focus levels $\hat{F}$. Instead of fitting the focus level to a specific distribution (e.g. Gaussian [32] or polynomial [37]), we determine the sub-frame depth stably by taking a weighted average based on the peak and 5 before and after focus level for each discrete depth.

5. Experimental Results

We experimentally validated our method using synthetic images and real microscopic polarization images.

5.1. Evaluation with Synthetic Data

We quantitatively evaluate the shapes recovered by our method and other methods using synthetic images. We render a polarization image of a diffuse object with an uneven and untextured surface using the pBRDF dataset [8] and synthesize a focal stack based on Eq. 8. The light source is positioned coaxially with the camera, and each scene has 30 focal stacks for objects of approximately 1mm in size. We add irregularities as a uniform distribution with a maximum of R_z mm to the depth of the mesh model (*Sphere*, *Bunny* and *Cheburashka*). We also add the zero-mean Gaussian noise with 0.02 standard deviation [38] to the raw input intensities ranging in $[0, 1]$, considering noise injection.

As baseline methods for comparison, we apply the method of Surh *et al.* [38], Won *et al.* [45] (trained their

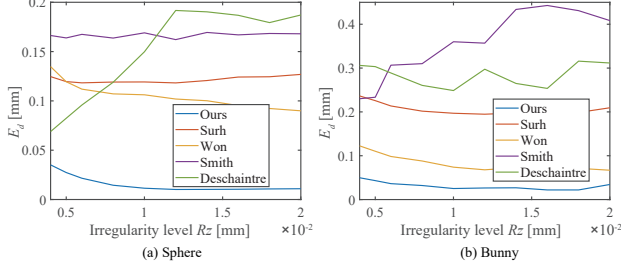


Figure 7. Error in different irregularity level. The lower irregularity level R_z corresponds to the smooth surfaces. Our method successfully recovers the shape regardless of the irregularity levels.

model on FlyingThings3D) as the SfF method and the method of Smith *et al.* [36], Ba *et al.* [6], and Deschaintre *et al.* [12] as the SfP method. In this paper, our proposed method consistently uses $(w_1, w_2) = (0.8, 0.6)$ for the contour direction and $(w_1, w_2) = (1.0, 0.8)$ for the other directions. Regarding the weight, the result becomes like a step if the weight difference is too significant. On the other hand, when the weight difference is too small, there is almost no change compared to the baseline BP. We empirically found this parameter. We input the stack $\mathfrak{S}_0$ for the SfF method, assuming a regular camera, and input a single all-in-focus polarization image for the SfP methods. To compare the surface normals, we calculate our method’s normals by creating meshes with screened Poisson surface reconstruction [22] with 6 octree depth, instead of relying on the DoPs.

Baseline comparisons Fig. 6 shows the results of shape reconstruction of surfaces with $R_z = 5.0 \times 10^{-3}$ mm irregularities. The values of error E_d and E_n represent RMSEs of the depth and the mean angular error (MAE) of the normal in degrees, respectively. The results show that our method achieves quantitatively the best accuracy in depth reconstruction. It clearly shows an advantage over the SfF method without polarization [38, 45] for textureless objects. In addition, although our method does not use DoP, its recovered normals are comparable to those of state-of-the-art SfP methods, and it can recover bumpy areas where other SfP methods [6, 36] smooth them out.

Comparison with different irregularity levels We quantitatively evaluate the reconstruction accuracy for different irregularity levels. We add different irregularities up to $R_z = 0.02$ mm to the mesh model of 1 mm in size. Fig. 7 shows the RMSEs of reconstructed depth for different R_z . In contrast to the irregularity tends to interfere with the accuracy of the single-shot SfPs, which assumes surface smoothness, our method effectively leverages different levels of irregularities.

Table 1. Ablation study.

		s_0	Stokes	Stokes+BP	Stokes+BP*
<i>Sphere</i>	E_d	0.072	0.052	0.030	0.027
	E_n	6.85	6.62	6.43	6.42
<i>Bunny</i>	E_d	0.14	0.090	0.048	0.043
	E_n	15.6	14.3	10.4	10.1
<i>Cheburashka</i>	E_d	0.12	0.076	0.051	0.047
	E_n	9.55	8.63	7.80	7.70

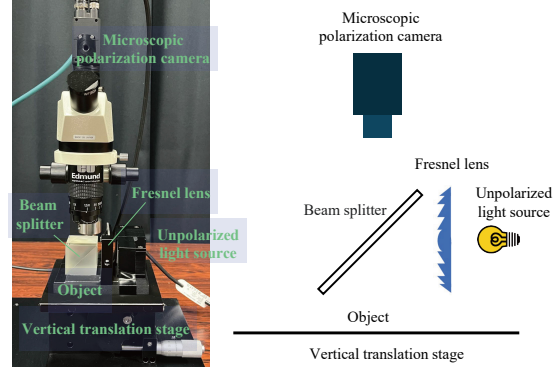


Figure 8. Experimental setup.

Ablation study We conduct an ablation study to quantitatively evaluate the effectiveness of our FM based on the Stokes vector (Sec. 4.1) and our BP (Sec. 4.2). Table 1 shows the result of E_d : the RMSE of the estimated depth and E_n : the MAE of the estimated normals for the same objects in Fig. 6. We compare the FM from regular camera images (“ s_0 ”), the FM from Stokes images using the polarization camera (“Stokes”), the FM with the BP assuming a simple weighting (“Stokes+BP”), and the FM with our BP (“Stokes+BP*”). For the joint weights of the simple BP (“Stokes+BP”), the weights $(w_1, w_2) = (1.0, 0.8)$ is used regardless of the adjacent direction. Also, the normals are calculated using the same method as ours. The results clearly show that our polarization FM utilizing the Stokes vector provides more accurate focus levels, and our BP based on the iso-depth contour concept allows for further better depth estimation.

5.2. Real-world Experiments

We also build a 3D sensing system with a microscope polarization camera. Fig. 8 shows an overview of our experimental setup. We replaced the camera of an off-the-shelf microscope camera system with a monochrome polarization camera (LUCID TRI050S1-PC) that uses quad-Bayer polarization filter chips (Sony IMX264MZR). We constructed coaxial falloff illumination by introducing a Fresnel lens and a non-polarizing beamsplitter for the side light source. We used fine micrometer movement to capture 30 focal stacks, moving the focal point at constant intervals. As in

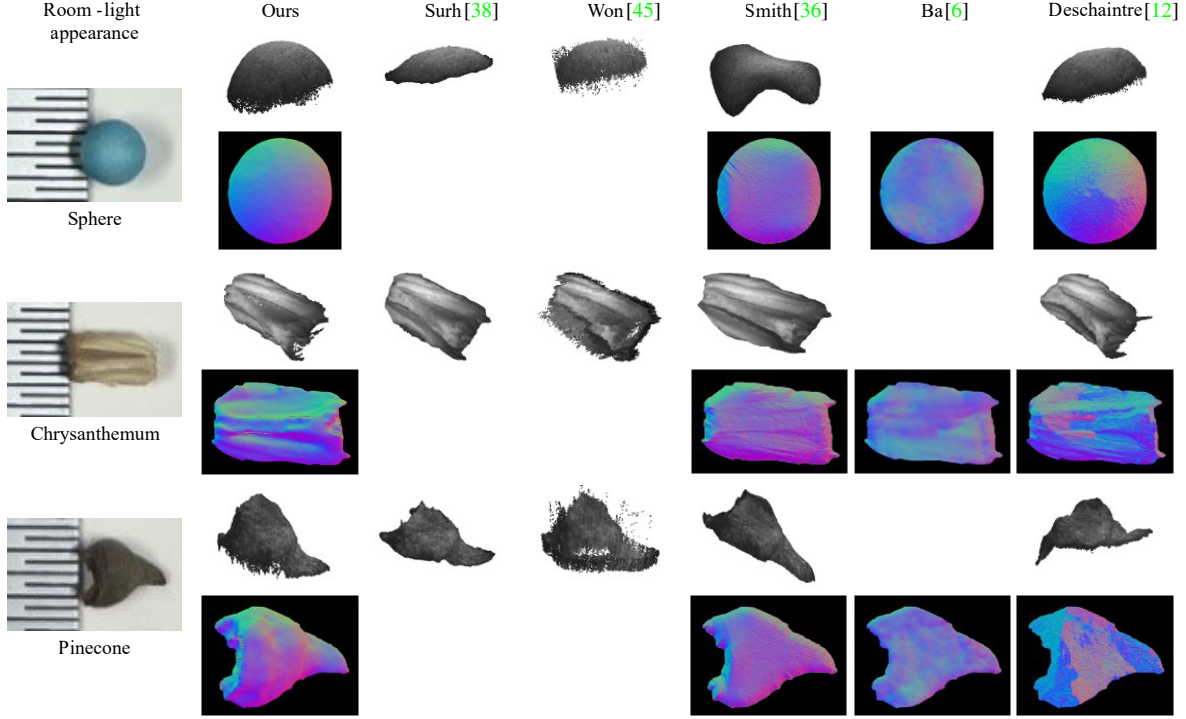


Figure 9. Result of reconstruction of real-world objects. The leftmost one shows the room light appearance (1 graduation on the ruler represents 0.5 mm). In each object, the upper row shows the 3D point cloud colored by the captured image, and the lower row shows the recovered normal. The blank area is because the corresponding method does not output the depths or the normals.

Sec. 5.1, we input the stack $\mathcal{S}_0$ for the baseline SfF method and input a single all-in-focus polarization image for the SfP methods for qualitative comparison.

Fig. 9 shows our results on textured or textureless real-world objects (a tiny *Sphere*, a *Chrysanthemum* seed, and a *Pinecone*). The objects are up to 3mm in size and have various shapes and surface irregularities. The results clearly demonstrate that our method successfully recovers the shape of these objects regardless of the texture or the irregularities of the surface. The results of the 3D point cloud and the recovered normals (the upper and the lower row for each object in Fig. 9) show that our method provides globally accurate shapes for all objects.

Limitations The major limitations of our method are as follows. Since our method relies on SfF, depth accuracy may be compromised when only a few focal stacks exist. One potential solution is to model a distribution of focus levels based on spatial frequency and incorporate it into the prior distribution of our BP in Sec. 4.2. However, that would not fundamentally address false peaks due to noise or FMs with finite window sizes. The other limitation is that our method relies on the polarization of diffuse reflections of opaque objects. Although our method does not rely on noisy diffuse DoPs, our BP does not work well for whole

black objects. The same happens to perfectly transparent objects containing complex optical paths or environment severely affected by diffraction. Strong indirect light can also affect especially the AoPs, which may degrade recovery performance in objects with strong inter-reflections.

6. Conclusion

In this paper, we introduced a novel Shape from Focus method utilizing polarization for per-pixel depth recovery of tiny objects regardless of the surface texture. We derive a blurring model of polarization images and show that the Stokes vector can be leveraged to estimate the focus level for each pixel. We further proposed a novel method to exploit the normal cues for updating the global shape based on belief propagation with the iso-depth contour concept. Experimental results show that our method can recover accurate per-pixel depth of various tiny objects regardless of the texture on the object surfaces. Our future work includes leveraging specular polarization components and analyzing surface properties such as refraction index, reflectance, and surface roughness in the context of microfacet normals.

Acknowledgements: This work was supported by JSPS KAKENHI Grant Numbers JP20H00612 and JP22K17914.

References

- [1] Combination of images with diverse focuses using the spatial frequency. *Inf. Fusion*, 2:169–176, 2001. 2
- [2] Muhammad Bilal Ahmad and Tae-Sun Choi. Application of three dimensional shape from image focus in lcd/tft displays manufacturing. *IEEE Trans. Consumer Electronics*, 53, 2007. 2
- [3] Emma Alexander, Qi Guo, Sanjeev Koppal, Steven Gortler, and Todd Zickler. Focal flow: Measuring distance and velocity with defocus and differential motion. In *ECCV*, pages 667–682. Springer, 2016. 2
- [4] Youngeun An, Gwangwon Kang, Il-Jung Kim, Hyunsook Chung, and Jong-An Park. Shape from focus through laplacian using 3d window. *FCN*, 2:46–50, 2008. 2
- [5] Gary A Atkinson and Edwin R Hancock. Recovery of surface orientation from diffuse polarization. *IEEE TIP*, 15(6):1653–1664, 2006. 2
- [6] Yunhao Ba, A. Gilbert, Franklin Wang, Jinfa Yang, Rui Chen, Yiqin Wang, Lei Yan, Boxin Shi, and Achuta Kadambi. Deep shape from polarization. In *ECCV*, pages 554–571, 2020. 2, 6, 7, 8
- [7] Seung-Hwan Baek, Daniel S Jeon, Xin Tong, and Min H Kim. Simultaneous acquisition of polarimetric svbrdf and normals. *ACM TOG*, 37(6):268, 2018. 2, 3
- [8] Seung-Hwan Baek, Tizian Zeltner, Hyun Jin Ku, Inseung Hwang, Xin Tong, Wenzel Jakob, and Min H. Kim. Image-based acquisition and modeling of polarimetric reflectance. *ACM TOG*, 2020. 6
- [9] Michael Chen, Zachary F Phillips, and Laura Waller. Quantitative differential phase contrast (dpc) microscopy with computational aberration correction. *Optics express*, 26(25):32888–32899, 2018. 2
- [10] Edward Collett. *Field guide to polarization*. Spie Bellingham, 2005. 4
- [11] Zhaopeng Cui, Jinwei Gu, Boxin Shi, Ping Tan, and Jan Kautz. Polarimetric multi-view stereo. In *CVPR*, pages 1558–1567, 2017. 2, 5
- [12] Valentin Deschaintre, Yiming Lin, and Abhijeet Ghosh. Deep polarization imaging for 3d shape and svbrdf acquisition. In *CVPR*, pages 15567–15576, 2021. 2, 6, 7, 8
- [13] Yoshiki Fukao, Ryo Kawahara, Shohei Nobuhara, and Ko Nishino. Polarimetric normal stereo. In *CVPR*, pages 682–690, 2021. 2
- [14] Jan-Mark Geusebroek, Frans Cornelissen, Arnold W.M. Smeulders, and Hugo Geerts. Robust autofocusing in microscopy. *Cytometry*, 39(1):1–9, 2000. 2
- [15] P. Grossmann. Depth from focus. *PR*, 5(1):63–69, 1987. 1, 2
- [16] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003. 2
- [17] Jui-Ting Huang, Chun-Hung Shen, See-May Phoong, and Homer Chen. Robust measure of image focus in the wavelet domain. In *Int. Sympo. Intell. Sig. Proc. Commu. Sys.*, pages 157–160, 2005. 2
- [18] Tianyu Huang, Haoang Li, Kejing He, Congying Sui, Bin Li, and Yun-Hui Liu. Learning accurate 3d shape based on stereo polarimetric imaging. In *CVPR*, pages 17287–17296, 2023. 2
- [19] Cong Phuoc Huynh, Antonio Robles-Kelly, and Edwin Hancock. Shape and refractive index recovery from single-view polarisation images. In *CVPR*, pages 1229–1236, 2010. 2, 3
- [20] Inseung Hwang, Daniel S Jeon, Adolfo Munoz, Diego Gutierrez, Xin Tong, and Min H Kim. Sparse ellipsometry: portable acquisition of polarimetric svbrdf and shape with unstructured flash photography. *ACM TOG*, 41(4):1–14, 2022. 2
- [21] Tomoki Ichikawa, Shohei Nobuhara, and Ko Nishino. Spiders: Structured polarization for invisible depth and reflectance sensing. *ArXiv*, abs/2312.04553, 2023. 2
- [22] Michael Kazhdan and Hugues Hoppe. Screened poisson surface reconstruction. *ACM TOG*, 32(3), jul 2013. 7
- [23] Alankar Kotwal, Anat Levin, and Ioannis Gkioulekas. Passive micron-scale time-of-flight with sunlight interferometry. In *CVPR*, pages 4139–4149, 2023. 2
- [24] Chenyang Lei, Chenyang Qi, Jiaxin Xie, Na Fan, Vladlen Koltun, and Qifeng Chen. Shape from polarization for complex scenes in the wild. In *CVPR*, pages 12622–12631, 2022. 2
- [25] Fengqiang Li, Florian Willomitzer, Prasanna Rangarajan, Mohit Gupta, Andreas Velten, and Oliver Cossairt. Sh-tof: Micro resolution time-of-flight imaging with superheterodyne interferometry. In *ICCP*, pages 1–10, 2018. 2
- [26] Zhongwei Li and YF Li. Microscopic photometric stereo: A dense microstructure 3d measurement method. In *ICRA*, pages 6009–6014, 2011. 2
- [27] Hanyue Lou, Mingui Teng, Yixin Yang, and Boxin Shi. All-in-focus imaging from event focal stack. In *CVPR*, pages 17366–17375, June 2023. 2
- [28] Wan-Chun Ma, Tim Hawkins, Pieter Peers, Charles-Felix Chabert, Malte Weiss, Paul E Debevec, et al. Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination. *Rendering Techniques*, 9(10):2, 2007. 2
- [29] Aamir Saeed Malik and Tae-Sun Choi. A novel algorithm for estimation of depth map using image focus for 3d shape recovery in the presence of noise. *PR*, 41:2200–2225, 2008. 2
- [30] Sofiane Mihoubi, Pierre-Jean Lapray, and Laurent Bigué. Survey of demosaicking methods for polarization filter array images. *Sensors (Basel, Switzerland)*, 18, 2018. 4
- [31] Daisuke Miyazaki, Robby T Tan, Kenji Hara, and Katsushi Ikeuchi. Polarization-based inverse rendering from a single view. In *ICCV*, pages 982–982, 2003. 2
- [32] Shree K Nayar and Yasuo Nakagawa. Shape from focus. *IEEE TPAMI*, 16(8):824–831, 1994. 1, 2, 6
- [33] Trung Thanh Ngo, H. Nagahara, and R. Taniguchi. Shape and light directions from shading and polarization. In *CVPR*, pages 2310–2318, jun 2015. 2
- [34] Alex Paul Pentland. A new sense for depth of field. *IEEE TPAMI*, 9:523–531, 1985. 2

- [35] Said Pertuz, Domenec Puig, and Miguel Angel Garcia. Analysis of focus measure operators for shape-from-focus. *PR*, 46(5):1415–1432, 2013. 2, 4
- [36] William AP Smith, Ravi Ramamoorthi, and Silvia Tozza. Linear depth estimation from an uncalibrated, monocular polarisation image. In *ECCV*, pages 109–125. Springer, 2016. 2, 6, 7, 8
- [37] M. Subbarao and Tao Choi. Accurate recovery of three-dimensional shape from image focus. *IEEE TPAMI*, 17(3):266–274, 1995. 6
- [38] Jaeheung Surh, Hae-Gon Jeon, Yunwon Park, Sunghoon Im, Hyowon Ha, and In-So Kweon. Noise robust depth from focus using a ring difference filter. In *CVPR*, pages 2444–2453, 2017. 2, 6, 7, 8
- [39] Trung Ngo Thanh, Hajime Nagahara, and Rin ichiro Taniguchi. Shape and light directions from shading and polarization. In *CVPR*, pages 2310–2318, 2015. 2
- [40] Andrea Thelen, Susanne Frey, Sven Hirsch, and Peter Hering. Improvements in shape-from-focus for holographic reconstructions with regard to focus operators, neighborhood-size, and height value interpolation. *IEEE TIP*, 18(1):151–157, 2009. 2
- [41] Chaoran Tian, Weihong Pan, Zimo Wang, Mao Mao, Guofeng Zhang, Hujun Bao, Ping Tan, and Zhaopeng Cui. Dps-net: Deep polarimetric stereo depth estimation. In *ICCV*, pages 3546–3556, 2023. 2
- [42] Lei Tian, Jingyan Wang, and Laura Waller. 3d differential phase-contrast microscopy with computational illumination using an led array. *Optics letters*, 39(5):1326–1329, 2014. 2
- [43] Silvia Tozza, William AP Smith, Dizhong Zhu, Ravi Ramamoorthi, and Edwin R Hancock. Linear differential constraints for photo-polarimetric height estimation. In *ICCV*, pages 2279–2287, 2017. 2
- [44] Silvia Tozza, Dizhong Zhu, William AP Smith, Ravi Ramamoorthi, and Edwin R Hancock. Uncalibrated, two source photo-polarimetric stereo. *IEEE TPAMI*, 44(9):5747–5760, 2021. 2
- [45] Changyeon Won and Hae-Gon Jeon. Learning depth from focus in the wild. In *ECCV*, pages 1–18. Springer, 2022. 6, 7, 8
- [46] Robert J Woodham. Photometric method for determining surface orientation from multiple images. *Optical engineering*, 19(1):139–144, 1980. 2
- [47] Hui Xie, Weibin Rong, and Lining Sun. Wavelet-based focus measure and 3-d surface reconstruction method for microscopy images. In *IROS*, pages 229–234, 2006. 2
- [48] Ge Yang and Bradley J Nelson. Wavelet-based autofocusing and unsupervised segmentation of microscopic images. In *IROS*, pages 2143–2148, 2003. 2, 4
- [49] Luwei Yang, Feitong Tan, Ao Li, Zhaopeng Cui, Yasutaka Furukawa, and Ping Tan. Polarimetric dense monocular slam. In *CVPR*, pages 3857–3866, 2018. 2
- [50] Jinyu Zhao, Yusuke Monno, and Masatoshi Okutomi. Polarimetric multi-view inverse rendering. In *ECCV*, pages 85–102. Springer, 2020. 2
- [51] Dizhong Zhu and William AP Smith. Depth from a polarisation+ rgb stereo pair. In *CVPR*, pages 7586–7595, 2019. 2